

Association for Information Systems AIS Electronic Library (AISeL)

PACIS 2007 Proceedings

Pacific Asia Conference on Information Systems
(PACIS)

2007

An Association between Bid Series and Transaction Price Facilitating Trading Decision Making Based on VPIP Algorithm

Zhe Zhang

City University of Hong Kong, janet.zhang@student.cityu.edu.hk

Jian Jiang

City University of Hong Kong, jianjiang5@student.cityu.edu.hk

Xiaoyan Liu

City University of Hong Kong, xiaoyan.liu@student.cityu.edu.hk

Huaiqing Wang

City University of Hong Kong, iswang@cityu.edu.hk

Follow this and additional works at: <http://aisel.aisnet.org/pacis2007>

Recommended Citation

Zhang, Zhe; Jiang, Jian; Liu, Xiaoyan; and Wang, Huaiqing, "An Association between Bid Series and Transaction Price Facilitating Trading Decision Making Based on VPIP Algorithm" (2007). *PACIS 2007 Proceedings*. 22.
<http://aisel.aisnet.org/pacis2007/22>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

54. An Association between Bid Series and Transaction Price Facilitating Trading Decision Making Based on VIPP Algorithm

Zhe Zhang

Department of Information Systems
City University of Hong Kong
janet.zhang@ student.cityu.edu.hk

Jian Jiang

Department of Information Systems
City University of Hong Kong
jianjiang5@ student.cityu.edu.hk

Xiaoyan Liu

Department of Information Systems
City University of Hong Kong
xiaoyan.liu@ student.cityu.edu.hk

Huaiqing Wang

Department of Information Systems
City University of Hong Kong
iswang@ cityu.edu.hk

Abstract

The stock price has long been an essential issue in stock market analysis, while rarely had any relevant study covered the relational effect of the bidder factor on stock price. Hereby we try to find an association between bid and transaction price time series. The fluctuation of pattern ratios embedded in the bid series affects the fluctuation of transaction price. That means when there is small changed amplitude of pattern occupancy ratio, there is a possibility that small fluctuation of transaction price will occur. And when there is a large changed amplify of pattern occupancy ratio, there is a possibility that large fluctuation of transaction price will occur. Also, we find the boundary value of changed amplify of pattern occupancy ratio which may lead to the relatively large fluctuation in transaction price. The finding contributes to provide reference information for the decision making of trading stock data. Additionally, we propose a new pattern-matching scheme called Visually and Practically Important Point (VIPP), which take the characteristic of the stock data into account. This scheme contributes that more valuable points can be remained in the stock time series after extracting the large number stock time series to small ones. Encouraging experiment is reported from the tests that there is an association between the bid sequence and transaction price time series of selected Chicago Stock Exchange, by using our pattern-matching scheme.

Keywords: Pattern Recognition, Visually and Practically Important Point, Decision Making

Introduction

Finding the factors which affect the stock price is a popular and important research problem (Fama 1965; Lendasse 2000). A popular method is to identify prototype patterns in the charts and forecast the future trend based on these chart patterns (Ge and Smyth, 2000). An even more interesting application of pattern matching in stock market time series data is to learn the frequently occurring patterns directly from the data, and build a model to forecast the stock market. Much work has done for finding the relationship between stock price and other factors for prediction. Such as Tetsuji (1992) used former transaction price for analysis, and Cheung et al. (2000) used the intra-day stock price to analysis and prediction. And Zhang (2005) try to find the relationship between the bid sequences. However, less attention has been paid to finding the relationship between the patterns in bid number sequence and the trading points, that is, whether the occurrence of trading price is related to certain patterns

embedded in the former bid number sequence. So the objective of this paper is to find the answer, in order to facilitating the decision making on stock trading.

On the other hand, there has been much work on defining subsequence-matching methods that try to find subsequences similar to a query sequence within a large time series (Moon et al. 2002). However, most of them did not consider the real meaning and characteristic of the data when doing pattern matching (Ge, 2000; Nesbitt and Barrass 2004). Regarding this, in additionally, we propose an algorithm called visually and practically important point (VPIP) pattern-matching scheme, which consider the characteristic of stock data to find the matched patterns for stock data mining. This scheme contributes that more valuable points can be remained in the stock time series after extracting the large number stock time series to small ones.

Encouraging experiment is reported from the tests that there is a relationship between the parameter, which is derived from the bid series defined by us, and the changing amplitude of trading price. Especially, when the parameter value exceeds a certain boundary level, the trading price will have a high probability to change. It contributes to provide reference information for the decision making of trading stock data.

Using our result found in this paper, we can find the possible fluctuation amplitude-changing trend and use it as the reliable base to make decisions. Thus, we can make an earlier decision on whether to transact stock or not. It can provide as a pre-noticing signal for those investors when large fluctuation in transaction price will occur.

The organization of this paper is as follows. In Section 2, we propose a new pattern-matching scheme based on Visually and Practically Important Points called VPIP. Section 3 introduces the design of each kinds of pattern style. Section 4 describes the design of the experiment. The result and analysis are reported in Section 5. And the final section addresses some conclusions of the paper.

Pattern Matching Scheme Based on Visually and Practically Important Points

To facilitate evolutionary stock time series segmentation, defining a similarity between time series for fitness evaluation is of fundamental important. In this section, we proposed a new time series pattern-matching scheme that was based on both the fact that frequently appearing pattern are typically characterized by a few critical point points, and the fact that the critical points have the practical meaning for the stock time series. For example, for the head-and-shoulder pattern, the turning points such as the highest and lowest points should be extracted, because these points are visually important in the identification process. In addition, those points that are near the transaction time should have the higher priority to be extracted, because they have more important effect on the transaction and more practically important for the stock time series. The proposed scheme follows the above idea by location the points that considering both the two aspects mentioned above.

Identification of Visually Important Points

Firstly, we need to identify the visually important points in the stock time series, which is similar but different with the algorithm proposed by Chung et al. (2004). The following scheme follows the principle of identifying the visually lowest and highest points in a data sequence $P = \{p_k | k = 1, \dots, m\}$, where m denotes the length of P , in accordance with a query

sequence $Q = \{Q_k | k = 1, \dots, n\}$, where n denotes the length of Q and $n \leq m$. The pseudocode is described in Figure 1.

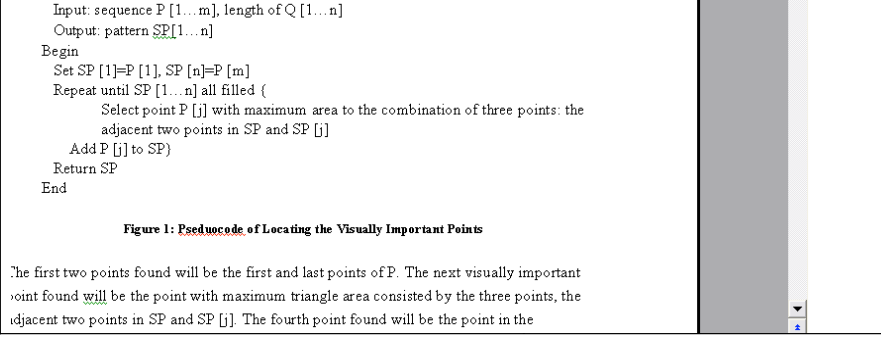


Figure 1: Pseudocode of Locating the Visually Important Points

The processes of locating the points described in Figure 1 are as the following. The first two points that are found will be the first and last points of P . The next visually important point that is found will be the point with maximum triangle area consisted by the three points, the adjacent two points in SP and $SP[j]$. The fourth point that is found will be the point in P with the maximum area that contained with the three points: first point, second point, and point between them, or contained with the three points: second point, last point, and the point between them. The process of extracting points will continue until the number is equal to the n . We can calculate the area with following formula by Helen theory.

$$S = \sqrt{l(l-a)(l-b)(l-c)} \quad (1)$$

$$l = (a + b + c) / 2 \quad (2)$$

where S is the area of triangle consisted by the three points mentioned above. In formula 2,

$$a = \sqrt{(y_3 - y_1)^2 + (x_3 - x_1)^2}, \quad b = \sqrt{(y_2 - y_3)^2 + (x_2 - x_3)^2} \quad \text{and} \quad c = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}$$

In the formula, a is the distance from P_3 to P_1 , which consist of one side of triangle area; b is the distance from P_2 to P_3 ; and c is the distance from P_1 to P_2 . Through this formula, we can extract the turning points that have the relatively highest or lowest value. To illustrate the identification process, the head-shoulder pattern is used in Figure 2 to show how to extract 4 points ($n=4$) from 7 points ($m=7$).

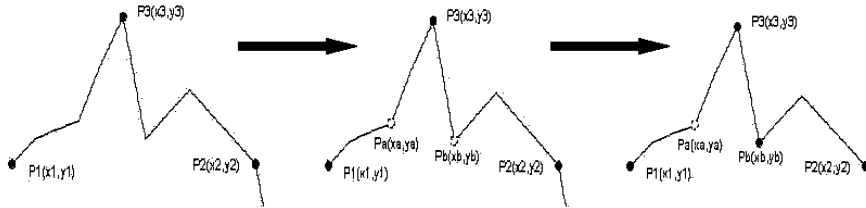


Figure 2: Identification of 5 Points Extracted from 7 Points

Identification of Practically Important Points

On the other hand, we should take the practical meaning of the stock data when extract the points from the time series. Those points that closer to the transaction time have the higher

priority than the far ones. In another word, the points have different distance to the transaction points. Therefore, we will propose to assign each of the points with a coefficient according to the distance between the point and the transaction point, the closer the distance, the higher weight the coefficient. Respectively, we calculate the distance λ from the following formulation, that is

$$\lambda = |x_{pt_b} - x_{pt_a}| \quad (3)$$

To illustrate the process of identification of these points, Figure 3 is used to show.

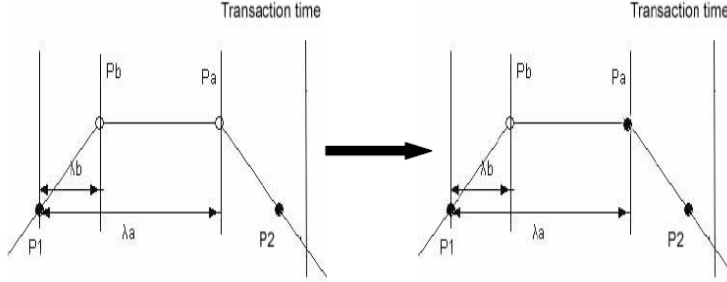


Figure 3: Identification of Practically Important Points

From Figure 3 we can find that the P_a has higher priority to be extracted than P_b for it is closer to the transaction point time and transaction.

Identification of Both Visually and Practically Important Points

Considering both sides of the above, we will propose the scheme that combines the two aspects mentioned in both Section 2.2 and Section 2.3. We will assign a coefficient, represented by ξ , that include both the area S and the distance λ . The coefficient is calculated by the following formulation.

$$\xi = \lambda_i \times S_i \quad 1 \leq i \leq m \quad (4)$$

We call it as Visually and Practically Important Points scheme (VIP). The pseudocode is described in Figure 4.

```

Function VPIPLocate (P, Q)
  Input: sequence P [1...m], length of Q [1...n]
  Output: pattern SP[1...n]
Begin
  Set SP [1]=P [1], SP [n]=P [m]
  Repeat until SP [1...n] all filled {
    Select point P [j] with maximum  $\xi$ 
    Add P [j] to SP
  }
  Return SP
End

```

Figure 4: Pseudocode of Locating the Visually and Practically Important Points




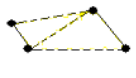
Pattern Representation

In order to obtain the objective of finding the relationship between the patterns in bid number sequence and the trading points, the first step is to find the matched patterns, each of which appear prior to the every trading point in the time domain. Developing a time series pattern

classification is of fundamental important for the pattern finding and matching. The general manner of design for the pattern classification can be characterized procedurally in the following general manner (Ge and Smyth 2000): (i) To find an approximation and robust representation for a time series; (ii) To define a flexible indexing function that can handle various pattern deformations; and (iii) To provide an efficient scalable and updatable algorithm by using the adopted representation and matching function for massive time series data sets. Several patterns with different numbers of PIP have been introduced. (Fu and Chung 2004). In this paper, we use several major stock patterns that occurred most frequently defined by Zhang et al. (2005), which are based the point number and the permutation and combination of slope (positive or negative) between two adjacent points.

Suppose that there are four points in sequence from the left to the right: Sp1, Sp2, Sp3, Sp4. Because there are three lines connected each two adjacent points, that is, line 1 connecting Sp1 and Sp2, line 2 connecting Sp2 and Sp3, and line 3 connecting Sp2 and Sp3. We just choose the leftmost line 1 and rightmost line 3 for the design because through this way the patterns can be roughly classified into four shape categories in Table 1, much simplifier and more convenient for analyze. Therefore, the combination of slope (positive or negative) can be illustrated as: $Slope(k_1, k_3) \in \{(+, -), (+, +), (-, +), (-, -)\}$, while k_1 and k_3 denote the slopes of line 1 and line 3 respectively. The suggested pattern names, the relationships among Sp1 to Sp4, and the pattern figures are shown in Table 1.

Table 1: Design of 4-point-Pattern

Pattern Name	Up-Trapeziform	Up Flag	Down-Trapeziform	Down Flag
Pattern Description	Sp1<Sp2 AND Sp3>Sp4	Sp1>Sp2 AND Sp3<Sp4	Sp1>Sp2 AND Sp3<Sp4	Sp1>Sp2 AND Sp3>Sp4
Pattern Figure				
	Sp1 Sp2 Sp3 Sp4 Up-trapeziform	Sp1 Sp2 Sp3 Sp4 Up flag	Sp1 Sp2 Sp3 Sp4 Down-trapeziform	Sp1 Sp2 Sp3 Sp4 Down flag

Experimental Design

The purpose of the experiment is to find the relationship between the changing ratio trends of each different four-point-pattern occurrence percentage, which appear prior to the trading points in the bid number sequence, and the change trend of transaction price. In the stock market, there are bid and ask price. The buyer states what price they will pay for the stock – this is the bid price. Under each bid price, there are sequences, which means the total numbers of how many shares are bided. For example, 300 shares are bided under price \$15.5. In order to predict the changing amplitude of transaction price based on the occupancy change ratio for each pattern, we try to find the rule that contained in the relationship between them.

The first step is to find the total number of each pattern between two transaction time.

Suppose that N_{t_i} is the matrix, which contain the number of four different patterns. It is calculated as the following:

$$N_{t_i} = \begin{bmatrix} N_{P1t_i-1} & N_{P2t_i-1} & N_{P3t_i-1} & N_{P4t_i-1} \end{bmatrix} \quad (5)$$

In (5), $N_{P_1 t_i t_{i-1}}$ is the number of Pattern 1 occur before the transaction time t_i , and between the two transaction time t_{i-1} and t_i . $N_{P_2 t_i t_{i-1}}$, $N_{P_3 t_i t_{i-1}}$ and $N_{P_4 t_i t_{i-1}}$ have the same meaning from Pattern 2 to Pattern 4.

In the second step, we will compute the occupancy rate of each pattern between transaction time t_{i-1} and t_i . The formula is shown as following. We use a matrix, P_{t_i} to represent the four different patterns' occupancy rate.

$$P_{t_i} = \begin{bmatrix} N_{P_1 t_i t_{i-1}} / \sum_{j=1}^4 N_{P_j t_i t_{i-1}} & N_{P_2 t_i t_{i-1}} / \sum_{j=1}^4 N_{P_j t_i t_{i-1}} & N_{P_3 t_i t_{i-1}} / \sum_{j=1}^4 N_{P_j t_i t_{i-1}} & N_{P_4 t_i t_{i-1}} / \sum_{j=1}^4 N_{P_j t_i t_{i-1}} \end{bmatrix}$$

$$(6) \quad = \begin{bmatrix} R_{P_1 t_i t_{i-1}} & R_{P_2 t_i t_{i-1}} & R_{P_3 t_i t_{i-1}} & R_{P_4 t_i t_{i-1}} \end{bmatrix}$$

So $R_{P_j t_i t_{i-1}}$, $j=1,2,3,4$ is the occupancy rate of j th pattern between the transaction time t_{i-1} and t_i .

In the third step, in order to find the relationship between changed the four-pattern occupancy ratio trend and the fluctuations of transaction price, we should define a parameter called the vector distance of two vectors, whose coordinates are regarded as four patterns' occupancy rate derived from the bid sequence, represented by Γ . It can be computed in (7). It will represent the changing amplitude of the four-pattern occupancy ratio. And on the other hand, we use K_i to represent the changing amplitude of transaction price at t_i , and use a vector distance Δ , between two transaction price, to represent the changing amplitude of transaction price between t_i and t_{i-1} . It can be computed in (8). Then we can find the relationship between Γ and Δ .

$$\Gamma = \left| \overrightarrow{P_{t_i} P_{t_{i+1}}} \right| = \left| (R_{P_1 t_i t_{i-1}}, R_{P_2 t_i t_{i-1}}, R_{P_3 t_i t_{i-1}}, R_{P_4 t_i t_{i-1}})(R_{P_1 t_{i+1} t_i}, R_{P_2 t_{i+1} t_i}, R_{P_3 t_{i+1} t_i}, R_{P_4 t_{i+1} t_i}) \right| \quad (7)$$

$$= \sqrt{(R_{P_1 t_{i+1} t_i} - R_{P_1 t_i t_{i-1}})^2 + (R_{P_2 t_{i+1} t_i} - R_{P_2 t_i t_{i-1}})^2 + (R_{P_3 t_{i+1} t_i} - R_{P_3 t_i t_{i-1}})^2 + (R_{P_4 t_{i+1} t_i} - R_{P_4 t_i t_{i-1}})^2}$$

$$\Delta = \left| \overrightarrow{K_i K_{i+1}} \right| = |K_{i+1} - K_i| \quad (8)$$

The used data set is taken from the time series of Chicago stock market. 10000 trading points are considered, together with the transaction price and their former bid time series.

Analysis of Experimental Results

In order to find the relationship between the Γ and Δ , firstly, we divide the Γ and Δ into different ranges. We divide the Γ into three ranges indicating small, middle and large changing amplitude of pattern occupation, respectively, from 0 to 5, 5 to 7 and 7 to 11.

Similarly, we classify the Δ into two kinds, which represent whether the amplitude of transaction price will change or not. Concretely, we divide the changing part into two kinds, the relatively small and large changing amplitude of transaction price. We choose 0 to indicate no change, 1 to indicate relatively small change and 2 to indicate relatively large change for Δ .

In the experiment, we try to exam the impact of amplitude changing in Γ on the amplitude changing of Δ from two aspects. Firstly, we control the Γ changed in a fixed range to find the percentage of how much transaction price will change due to different ranges. The result can

be seen in Table.2. From the table, we can find that the larger value of Γ will affect the larger Δ . For example, when Γ is in the small range from 0 to 5, there is relatively small percentage of Δ change, 49.38%. And when Γ is in a large range from 7 to 11, there is a relatively large percentage of Δ change, 76.0%. Also, we find there is a boundary value of Γ for the price amplitude that will have obvious change, when $\Gamma = 7$. Also, comparing the two occupation to the changing situation column, we find that the relatively small Γ is related to the relatively small Δ and large Γ is related to the relatively large Δ . For example, when Γ is in the relatively small range from 0 to 5, there are relatively higher possibilities for the occurrence of Δ in the small range than that in the high range where Γ is from 7 to 11, comparing 75.01% and 57.89%. Similarly, when Γ is in the relatively high range from 7 to 11, there is a higher possibility for the occurrence of Δ in the large range than that in the small range where Γ is from 0 to 5, comparing 42.11% and 24.99%. In a word, we can find that if there is a higher value of Γ , there is a higher possibility that larger amplitude change of transaction price will occur.

Table 2: The Probability of Price-Changing Rate under Fixed Γ Range

	$\Delta=0$	$\Delta>0$		
	$\Delta=0$ (%)	$\Delta=1$ /Occupation to the changing situation (%)	$\Delta \geq 2$ /Occupation to the change situation (%)	$(\Delta=1)+(\Delta \geq 2)$
$0 \leq \Gamma < 5$	50.62	37.04/75.01	12.34/24.99	49.38%
$5 \leq \Gamma < 7$	41.38	41.38/70.59	17.24/29.41	58.62%
$7 \leq \Gamma < 11$	24.0	44.0/57.89	32.0/42.11	76.0%

Secondly, we control the Δ , to find the percentage of how many Γ will occur in a range due to whether the transaction price will change or not. The result is shown in Table.3.

Table 3: The Probability of Γ Occurrence Rate under Fixed Price Changing

	$0 \leq \Gamma < 5$ (%)	$5 \leq \Gamma < 7$ (%)	$\Gamma \geq 7$ (%)	$(5 \leq \Gamma < 7) + (\Gamma \geq 7)$ (%)
$\Delta=0$	43.64	35.67	20.69	56.36
$\Delta=1$	30.26	33.80	35.94	69.74
$\Delta \geq 2$	23.80	29.92	46.27	76.20

From Table.3, we can find that when Δ is relatively small, most occurrence percentage of Γ is located in the small range of Γ , seeing 43.64% is higher than 35.67% and 20.69%. And when Δ is relatively high (e.g. $\Delta \geq 2$), most occurrence percentage of Γ is located in the high range of Γ , seeing 46.27% is higher than 29.92% and 23.80%. In a word, the situation that transaction price is changed with high amplitude are mostly occur when Γ is in a high range. And situations that transaction price is changed with low amplitude are mostly occur when Γ is in a low range. This is coincident with the findings mentioned before.

Conclusion

From the experiment result, we find an association between bid series and transaction price. There is a relationship between the changed amplitude of pattern occupancy ratio derived from bid sequence and the fluctuation amplitude of transaction price, that is, when there is the small changed amplitude of pattern occupancy ratio, there is a possibility that small fluctuation of transaction price will occur. And when there is a large changed amplitude of

pattern occupancy ratio, there is a possibility that large fluctuation of transaction price will occur. Also, we find the boundary value of changed amplify of pattern occupancy ratio which may affect the relatively large fluctuation in transaction price. The result contributes to provide reference information for the decision making of trading stock data.

On the other hand, we propose a new pattern-matching scheme called Visually a Practically Important Point (VIP), which consider more characteristic of stock data than other schemes when finding the matched patterns for stock data mining. This scheme contributes that more valuable points can be remained in the stock time series after extracting the large number stock time series to small ones.

References

- Campbell J., Lo, A., and MacKinlay A. C. *The Econometrics of Financial Markets*, Princeton University Press, 1997.
- Cheung, W.S., Ng, H.S., and Lam, K.P. "Intraday stock price analysis and prediction," *Proceedings of the 2000 IEEE International Conference*, 2000, pp. 47-52.
- Chung, F. L., Fu, T. C., Ng, V. and Luk, R. "An Evolutionary Approach to Pattern-based Time Series Segmentation," *Evolutionary Computation, IEEE Transactions*, (8:5), Oct. 2004, pp. 471 – 489.
- Fama E. "The Behavior of Stock Market Prices," *Journal of Business*, 38, 1965, pp. 34-105.
- Fu, T. C., Chung, F. L., Luk, R. and Ng, V. "Flexible Time Series Pattern Matching Based on Perceptually Important Points," *International Joint Conference on Artificial Intelligence Workshop on Learning from Temporal and Spatial Data*, 2001, pp. 1-7.
- Fu, T. C., Chung, F. L., Luk, R., and Ng, C. M. "Financial Time Series Indexing Based on Low Resolution Clustering," *Proceedings of The 4th IEEE International Conference on Data Mining (ICDM-2004)*, Brighton, UK, 2004, pp. 5-14.
- Ge, X., and Smyth, P. "Deformable Markov Model Templates for Time-Series Pattern Matching," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston, MA, 2000, pp. 81-90.
- Lendasse A., De B. E., Wertz V. and Verleysen M. "Nonlinear Financial Time Series Forecasting on Application to Bel20 Stock Market Index," *European Journal of Economic and Social Systems* (14:1), pp 81-91, 2000.
- Moon, Y. S., Whang K.Y., and Han W.S. "General Match: a Subsequence Matching Method in Time-series Databases Based on Generalized Windows," *SIGMOD*, 2002, pp. 382–393.
- Nesbitt, K.V., and Barrass, S. "Finding Trading Patterns in Stock Market Data," *IEEE Computer Graphics and Application*, 2004, pp. 45-55.
- Tanigawa, T., and Kamijo, K. "Stock price pattern matching system-dynamic programming neural networks approach," *Neural Networks*, 1992, pp. 465-471.
- Zhang, Z., Liu, X. Y., and Wang, H. Q. "Discovery in Stock Time Series Based on Perceptually Important Point Algorithm," *Proceedings of San Diego International Systems Conference*, San Diego, LA, USA, 2006.
- Zhang, Z., Li, J., Wang, S., Wang, H. Q. "Study of Principal Component Analysis on Multi-dimension Stock Data," *Chinese Journal of Scientific Instrument*, (26: 8), 2005, pp. 2489-2491.